(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 522 868 A1

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
13.04.2005 Bulletin 2005/15

(51) Int Cl.7: **G01S 5/18**, G01S 3/808

(21) Application number: **03022983.5**

(22) Date of filing: **10.10.2003**

(84) Designated Contracting States:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IT LI LU MC NL PT RO SE SI SK TR
Designated Extension States:
AL LT LV MK

(71) Applicant: **Harman/Becker Automotive Systems GmbH**
76307 Karlsbad (DE)

(72) Inventors:
- **Hess, Wolfgang Georg**
  58452 Witten (DE)
- **Nitzpon, Hans-Jürgen**
  76337 Waldbronn (DE)
- **Zeller, Michael**
  76307 Karlsbad-Spielberg (DE)

(74) Representative: **Beetz & Partner Patentanwälte**
Steinsdorfstrasse 10
80538 München (DE)

(54) **System for determining the position of a sound source and method therefor**

(57) Method for determining the position of a sound source (11) in relation to a reference position, comprising the steps of generating a sound signal emitted from the sound source (11), detecting the emitted sound signal, processing the sound signal by the use of a physiological model of the ear, deducing at least one of lateral deviation in relation to the reference position, time delay of the sound signal from the sound source to the reference position, and the sound level of the detected sound signal.

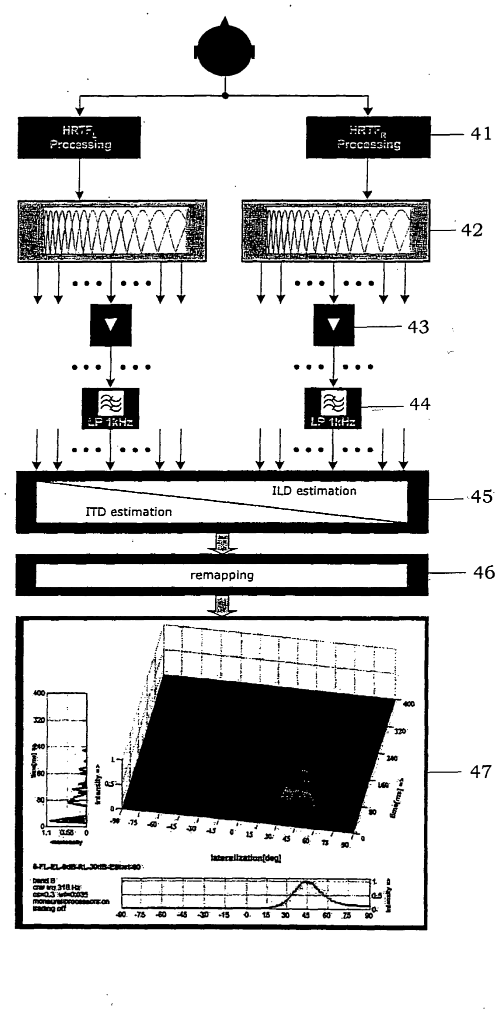


Fig. 4

EP 1 522 868 A1

Printed by Jouve, 75001 PARIS (FR)

Description

[0001] The invention relates to a system and a method for determining the position of a sound source in relation to a reference position. The invention relates especially to a method for determining the individual positions of a plurality of sound sources of a multi channel audio system.

Related Art

[0002] Recently, multi channel audio and video systems have become more and more popular, especially for home entertainment application. In these systems the sound is stored on a storage medium such as a DVD or a SACD (super audio CD) and the sound is encoded in bit streams for each channel of the multi channel audio system. Normally, home entertainment audio or video systems comprise five loudspeakers and one subwoofer. In order to optimize the sound produced from this loudspeaker system, the loudspeakers should have a special geometrical arrangement relative to each other. In the case of a 5.1 surround setup comprising five loudspeakers there exist standards (e. g. ITU-R BS 775-1 for 5.1 systems). In these standards the position of the loudspeakers relative to an optimal listening position - called sweet-spot is defined. By way of example, in a 5.1 surround setup the five loudspeakers should have the following position in an horizontal plane: One loudspeaker in the center at 0 ° azimuth in the horizontal plane, one front left loudspeaker at 30 °, one front right loudspeaker at -30 °, one left surround loudspeaker at 110 ° and one right surround loudspeaker at -110 °. This speaker setup will be explained in detail below. If these loudspeaker positions can be obtained the surround sound is optimal, as the sound was recorded and the different channels of the multi channel audio system were encoded in such a way that for the above-mentioned speaker setup the surround sound is optimal.

[0003] Especially, in home cinema applications, but also in professional applications within a professional sound studio it is not always possible to place the different loudspeakers at their ideal position. Furthermore, all loudspeakers should have the same distance relative to the listening position. This condition can also be difficult to fulfill, especially in the case where the loudspeakers have to be placed in a living room of an individual user of the system.

[0004] Available systems nowadays often comprise possibilities for adjusting the system to the room in which it is arranged. Therefore, the user of the system has to measure manually the distance of the different loudspeakers to the sweet-spot, the user has to determine the sound level of each loudspeaker at the sweet-spot and he or she has to determine the lateral position of deviation of the different loudspeakers. Especially, the determination of the angle of the loudspeaker relative to the center axis may cause difficulties.

[0005] Therefore, a need exists to automatically determine the position of a sound source such as a loudspeaker relative to a reference position.

Summary of the invention

[0006] This need is accomplished by the independent claims. In the sub claims preferred embodiments of the invention are described.

[0007] According to a first aspect of the invention the method for determining the position of a sound source in relation to a reference position comprises the following steps: First of all, a sound signal is generated, that is emitted from the sound source. This sound signal as emitted from the sound source is detected and the sound signal is processed by the use of a physiological model of the ear. Furthermore, at least one of lateral deviation in relation to the reference position, time delay of the sound signal from the sound source to the reference position and the sound level of the detected sound signal can be deduced. According to an aspect of the invention, the method for determining the position, time delay or sound level is based on a simulation of the human auditory signal processing. Using this simulation of the human auditory signal processing it is possible to automatically determine the parameters needed to obtain the position of a sound source such as a loudspeaker. This is especially helpful for the determination of the correct position of the loudspeakers in a multi channel audio system. With the method according to the invention the position, *i. e.*, the lateral deviation in relation to the reference position, the time delay and the sound level can be determined for each sound source separately. With the results obtained by the described method of the invention the different loudspeakers can be adapted to their ideal position as proposed by different standards.

[0008] According to a preferred embodiment the sound signal is detected by a sound receiving unit having two microphones arranged spatially apart resulting in two audio channels which are processed by the physiological model. According to a preferred embodiment, the two microphones are positioned in the right and left ear of a dummy with replication of the external ear. The auditory system can be described as a system with two input channels. To simulate binaural auditory events, several model exist which are usually divided into two groups of algorithms: One group is formed from physiologically driven models, the other one from psycho-acoustically driven approaches. For the purpose of evaluating acoustical properties of rooms, it is not meaningful to simulate every part of the human auditory system. It is sufficient to simulate an auditory pathway just accurately enough to be able to predict a number of psycho-acoustical phenomena which are of interest. Since the position of auditory event - localization seems to be fairly stable even in reverberant environments the invention developed the idea of evaluating audio set up configurations by a computational binaural model of auditory localization which is fed with the binaural room-impulse responses captured by a measurement method which is as close as possible to the way the human auditory system receives acoustic input. Therefore, according to the invention the sound signal emitted from the sound source is an impulse and the head-related, binaural impulse responses are detected.

[0009] The sound signal is processed by using a physiological model of the ear. This means that the path of the sound from the outer ear to the inner part of the head has to be simulated. Due to the fact that the two microphones of the sound receiving unit are positioned in the right or left ear of a dummy with replication of the external ear the filtering by the outer ears is already encoded in the measured head related impulse responses, i. e., the ear input signals. It should be understood that the microphones need not to be positioned in the ear of the dummy head. If no dummy head is used, however, the signal path from the outer ear to the inner ear has to be considered when simulating the path of the sound signal. The use of a dummy head with a detailed replication of the human outer ear helps to improve the results obtained by the simulation.

[0010] According to a preferred embodiment the sound signal of the left and the right ear is processed by applying a gammatone-filter bank to the recorded signal. This signal processing step simulates the inner ear. By the gammatone-filter bank the frequency deposition as performed by the cochlea is simulated.

[0011] According to a further aspect of the invention the sound signal is processed by a half-wave rectification of the recorded sound signal. This half-wave rectification of the sound signal is performed in order to mimic the firing behavior of the inner hair cell response. According to a further aspect of the invention the sound signal is processed by low-pass filtering the recorded sound signal to consider that the fine structure of the rectified signals cannot be resolved in the higher frequency bands and the signals are low-pass filtered after the half-wave rectification.

[0012] According to a further aspect of the invention the two channel audio signal is processed by carrying out a cross-correlation analysis of the sound signals of the left and the right ear, said cross-correlation analysis resulting in the interaural time differences (ITD) of the left and right ear. The interaural time difference (ITD) describes the time delay of a sound event that is received at two different points of time in the right and left ear depending on the distance the sound has to travel to the two ears.

[0013] According to a further aspect of the invention the signal is processed by introducing inhibition elements resulting in the interaural level differences (ILD) of the left and the right ear. The ILD describes the differences of sound levels of a sound that is received in the left and right ear. Depending on the position of the head relative to the sound source the sound level will differ on the right and the left ear, resulting in an interaural level difference (ILD).

[0014] According to a further aspect of the invention the signal processing comprises the step of deducing a binaural activity pattern (BAP) from the interaural time differences (ITD) and the interaural level differences (ILD), the binaural activity pattern comprising the intensity of the sound signal in dependence of time and comprising the intensity of the sound signal in dependence of the lateral deviation of the sound source. A transformation from an ITD-scale to a scale representing the position on the left-right deviation scale helps to determine the relative position of the sound source. The binaural activity pattern comprising the intensity of the sound signal in dependence of time and in dependence of the lateral deviation of the sound source allows the determination of the time delay, the determination of the intensity of the sound signal and the determination of the sound level. This is possible as the time delay can be deduced from the intensity of the sound signal in dependence of time, the lateral deviation can be deduced from the intensity of the sound signal in dependence of the lateral position of the sound signal relative to the reference position, and the sound level can be deduced from the maximum of the sound signal. As mentioned above, these three parameters (lateral position, sound level, and delay time) are necessary in order to determine the relative arrangement of the sound sources. According to the invention, the method provides as an output these three parameters so that the calculated positions and the sound levels can be used in order to correct the positions in accordance with a predetermined standard configuration such as the ITU-R BS.775-1 standard.

[0015] Instead of manually correcting the position of the sound sources to the position proposed by the standard after having calculated the actual position by the above described method it is also possible to provide the above calculated values to a processing unit or a manipulation unit of the audio system or to the processing unit of an external computer, so that the different channels comprising the sound of each sound source can be corrected in such a way that the emitted sounds corresponds to the sound that would have been produced by a loudspeaker system, if the loudspeakers were positioned as proposed by the standard. Therefore, the lateral deviation, the time delay and the sound level can be used in order to calculate correction data which adjust the sound signal emitted from the sound sources according to an arrangement of sound sources which correspond to a predetermined standard arrangement of the sound sources. This predetermined standard arrangement of the sound sources may correspond to the arrangement of the sound sources according to an international standard, be it the above-mentioned ITU standard or any other standard. After having determined the parameters (lateral deviation, time delay and the sound level) the different channels of the sound sources can be controled in such a way that an optimal sound is produced at the sweet-spot even, if the sound sources are not positioned at an ideal position suggested by the standards. The above-described methods for determining the position of a sound source is therefore especially dedicated for determining the individual positions of a plurality of sound sources of a multi channel audio system.

[0016] According to the invention the following system can be used for determining the position of a sound source. First of all, the system may comprise a sound generator for generating sound. The generated sound can be a sound signal in the form of an impulse which is emitted from a sound source of the system for emitting the generated sound. The system further comprises a sound receiving unit for detecting the sound emitted from the sound sources and sound analyzing means for processing the detected sound signal by the use of a physiological model of the ear and for deducing at least one of lateral deviation in relation to a reference position, time delay of the sound signal from the sound source to the reference position, and the sound level of the detected sound signal. The sound analyzing means processes the detected sound signal in such a way using the physiological model of the ear that the lateral deviation in relation to a reference position, or the lateral position relative to a predetermined axis, the time delay and/or the sound level can be determined. From these parameters the position of the sound source (the loudspeaker) relative to a listing position can be determined.

[0017] According to a preferred embodiment the sound receiving unit comprises two microphones which are arranged spatially apart. These two microphones simulate the receiving of the sound in the human ear so that the system further comprises a dummy with two ears in which the microphones are positioned, the ears comprising replications of the external ear. In order to simulate the auditory pathway as accurately as possible the ears comprise replications of the external ear so that the sound detected at the inside corresponds to the sound detected by the inner human ear.

[0018] The detected sound is then analyzed by the sound analyzing means. According to a preferred embodiment of the invention the sound analyzing means comprises a gammatone-filter bank, a half wave rectification unit, a low pass filter, and a cross-correlation analysis unit. The gammatone-filter bank, the half-wave rectification unit and low pass filter, through which the detected signal of the right and the left ear paths simulate the inner ear of human beings. Furthermore, filtering devices for filtering the outer ears are not necessary as the filtering by the outer ears is already encoded in the measured head-related impulse responses, i. e., the ear input signals. After the signals have passed the gammatone-filter bank, the half-wave rectification unit and the low pass filter, the signals pass a cross-correlation analysis unit in which the signals are transformed to ITD signals. Furthermore, the sound analyzing means comprises an inhibition unit comprising the inhibition function for determining the ILDs. From the ILDs and the ITDs the binaural activity pattern can be calculated. This binaural activity pattern of the head-related impulse responses can be used for the determination of position deviation, time delay and sound level, so that in case of several sound sources the system can determine the exact position of each sound source relative to a reference position and relative to the other sound sources.

[0019] The system may further comprise a manipulation unit for manipulating the sound emitted by the sound source for compensating a mismatch of actual position of the different sound sources relative to the reference position of each sound source. This manipulation unit can be used to adjust a non-optimal geometrical arrangement of the sound sources (loudspeakers) by an algorithm which uses the detected delay time, sound level and position deviation. Using these three parameters the manipulation unit can manipulate the different channels of a multi channel audio system in such a way that an ideal sound can be obtained even though the loudspeakers are not positioned as proposed by the standard.

[0020] Other features, objects and advantages of the present invention will become apparent in the course of the following description, referring to the accompanying drawings.

Fig. 1 is a schematic view of a system for determining the position of a sound source,
Fig. 2 shows the arrangement of a multi channel audio system according to the ITU-RBS.775-1 standard,
Fig. 3 shows a head-related coordinate system,
Fig. 4 shows a binaural model for calculating the position of the sound source,
Fig. 5 shows a cross-correlation function inhibition structure used for determining the ITDs and ILDs,
Fig. 6 shows a flowchart for determining the position related parameters of a sound source, and
Fig. 7 shows in further detail a signal processing of the sound signal for determining the position related parameters.

[0021] Fig. 1 shows a system 10 for determining automatically the position of sound sources 11. In the embodiment shown, five sound sources or loudspeakers 11 are shown. These five loudspeakers could be part of a multi channel audio or multi channel video system of a home entertainment system. They could also be part of a professional audio system in a sound studio or part of any other system incorporating sound sources. For detecting the sound emitted from the different sound sources 11 a dummy head 12 with a replication of the human ear is arranged in the position, the user of the sound system wants to use as listening position. The signals emitted from the sound sources 11 are picked up by the dummy head (e. g., a Neumann KU-80 with detailed replications of the external ear). The ears carry microphones 13 in the ear channels (e. g., Sennheiser KE-4 microphones). The sound detected from the microphones

13 is then transmitted to a computer unit 14. This computer unit 14 can be a normal personal computer and comprises a sound analyzing unit 15 in the sound manipulation unit 16. It should be understood that the computer unit 14 may further comprise a soundcard (not shown) for generating the sound emitted from the loudspeakers 11. Furthermore, the computer unit 14 may comprise a display unit for displaying the calculated processed audio signals. The five loudspeakers 11 may be installed in a living room (not shown) of an individual, in this kind of room the position of the different loudspeakers 11 relative to the listening position being different. In order to obtain an optimal surround sound for a surround sound setup the loudspeakers 11 have to be positioned at exact predetermined positions, and the distance of each loudspeaker from the listening position should be the same for each loudspeaker.

[0022] Fig. 2 shows an arrangement of loudspeakers as described in the ITU-R BS.775-1 standard. In this arrangement, all loudspeakers 11 should have the same distance D relative to a reference listening position 20 in the middle of the loudspeakers. This reference listening position 20 is also called sweet-spot. Furthermore, the worst case listening positions 21 are shown. These worst case listening positions are arranged at half of the loudspeaker base width B. In this ideal arrangement one center loudspeaker C is arranged in the middle of an axis 22 defined by the middle of the loudspeaker C and the reference listening position 20. One loudspeaker R is arranged on the front right side by 30 °, another loudspeaker L is arranged at -30 °, the right surround loudspeaker is arranged at 110 ° and the left surround loudspeaker is arranged at -110 °, all loudspeakers having the same distance D = B to the reference listening position. With this arrangement of the loudspeakers an optimal surround sound can be obtained, as the sound emitted from the sound sources was produced for exact this loudspeaker arrangement. As can be seen from Fig. 2 it can be very difficult to keep these predescribed positions of the loudspeakers in a room. First of all, it can be possible that the arrangement of the different loudspeakers differs from the arrangement in Fig. 1 in such a way, that the loudspeakers are arranged at a different distance, and/or the loudspeakers cannot be positioned at the exact angles. The manual determination of the exact position of the loudspeakers can be very difficult, especially the determination of the different angles. If the condition shown in Fig. 2 is not met, the surround sound emitted from the five loudspeakers may not be optimized. All loudspeakers 11 shown in Fig. 2 should not differ more than 7 ° in elevation and all loudspeakers should have the same sound level. For each loudspeaker three parameters are important, *i. e.*, the azimuth angle of every loudspeaker relative to the sweet-spot, the sound level of each loudspeaker and the delay time, if the loudspeakers cannot be arranged at the same distance to the sweet-spot.

[0023] For the determination of these three parameters a model of the human auditory system is used. In Fig. 3 the coordinate system used in this simulation is shown. The head or the dummy head is arranged in the middle of the system, the front direction lies in the direction of $\varphi = 0°$, $\delta = 0°$. In the figure shown in Fig. 3 φ represents the azimuth angle in the system and δ represents the elevation angle. Furthermore, the horizontal plane 31, the median plane 32 and the front plane 33 are shown. Using the coordinate system shown in Fig. 3 and the dummy head the binaural room impulse responses were measured. Further details for the coordinate system shown in Fig. 3 can be found in Blauert, J.: Spatial Hearing - the psychophysics of human sound localization (2nd enhanced edition), MIT Press, Cambridge, MA, 1997.

[0024] In Fig. 4 the different blocks for simulating a human ear are shown in more detail. The sound signal emitted from a sound source and detected by a human being can be simulated by the following signal path. First of all, the outer ear has to be simulated as sound signal is first received by the outer ear. The outer ears (pinnae) are asymmetrical (as to left and right ear) and can be described by head-related transfer functions (HRTF) 41 of the right and of the left ear. For simulating a complete signal path each HRTF function has to be considered. In the present case, however, these outer ear processing step can be omitted as a dummy head with detailed replication of the human outer ear was used and the microphones were positioned in the inner ear. This means, that the filtering by the outer ears is already encoded in the measured head-related ear input signals. The signal is then processed by simulating the inner ear (cochlea and hair cells), which in the flowchart shown in Fig. 4 corresponds to the blocks 42, 43, and 44. In order to simulate the frequency decomposition as performed by the cochlea, the left and the right ear signals are processed by a gammatone-filter bank 42. By way of example, 36 bands with a sample rate of 48 kHz, covering a frequency range from 23 to 21164 Hz (mid frequencies of the filters) can be used. To mimic the firing behavior of the inner-hair-cell response a simple half-wave rectification 43 is implemented. To consider that the fine structure of the rectified signals cannot be resolved in the higher frequency bands, the signals are low-pass filtered after the half-wave rectification in a low pass filter 44 (1st-order low pass, f_c = 1 kHz). Afterwards, a running cross-correlation analysis comparing the signal of the left and the right channel is carried out in each frequency band in step 45. The determination of the ILD and the ITD are described in more detail in reference to Fig. 5. To consider the sluggishness of the human binaural system, the temporal window in the running cross-correlation analysis was set to a duration of 10 ms (triangular window). The outputs of each frequency band are transformed or remapped from an ITD-scale to a scale representing the position on a left-right deviation scale (box 47). This scale represents one of the three coordinates of the binaural activity pattern to be rendered. For this scale transformation, the frequency-dependent relationship of ITDs (and ILDs) and the sideway deviation is determined from a catalog of head-related transfer functions HRTFs that was established using the same dummy head which was used to measure the binaural room impulse responses. The catalogue has a resolution of 5 ° azimuth in the horizontal plane and 10 ° elevation in the median plane, so that altogether 1163 directions were measured.

[0025] Multiple cross-correlation peaks as caused by the periodicity of the signal which are typically found at the contra-lateral side are eliminated by implementing contra-lateral inhibition into the cross-correlation algorithm. The interaural level differences are processed separately using excitation-inhibition cells. Here too, the outputs are transformed to an appropriate left-right-deviation scale.

[0026] Finally, the binaural activity pattern consists of the weighted output of the cross-correlation algorithm and the algorithm based on excitation-inhibition cells. This output is computed and displayed for the frequency band under consideration, thus leading to the three-dimensional pattern as shown in Fig. 4. The three-dimensional binaural activity pattern is shown in a three-dimensional scale in the box with reference numeral 47. The remapping is indicated in box 46. The graph on the bottom analyzes the intensity over a lateralization, i. e., the lateral deviation relative to the axis 22 shown in Fig. 2. From the graph at the bottom it can be deduced that the lateral deviation is approximately 45 °, so that the sound source is positioned 45 ° relative to the middle axis. The graph on the left side shows the intensity over time. From this graph the time delay of the sound signal can be deduced. As can be seen from the peak in the left hand graph the first direct sound arrives approximately after 20 ms, so that the distance between the sound source and the listening position can be determined when the sound propagation velocity is known. From the graph on the left hand side the sound level can also be determined from the peak intensity.

[0027] In conclusion, the above-mentioned physiological model of the ear and the corresponding signal processing can be used in order to determine the three parameters needed for the exact determination of the position of the sound sources, the parameters being the sound level, the lateral deviation and the time delay. With the use of a dummy head, two microphones and a normal personal computer, the positions of the sound sources relative to each other can be determined. If the determined positions do not correspond to the positions as prescribed by the standard configuration the sound sources can be repositioned and the method can be repeated. As will be described later on, these three parameters can also be used for manipulating the different channels of the sound system in such a way that the emitted sound will be detected at the listening position as if the sound sources were positioned at positions according to the standard.

[0028] In the following, the generation of the binaural activity patterns are described in more detail.

[0029] The binaural model algorithm is signal driven, *i. e.*, it has a strict bottom-up architecture. Its basic version had been developed by Lindemann (see Lindemann, W. (1982): Evaluation of interaural signal differences, in: O.S. Pedersen und T. Pausen (Hrsg.), Binaural effects in normal and impaired hearing, 10^{th} Danavox Symposium, Scandinavian Audiology Suppl. 15, 147-155). This algorithm was originally prompted by an attempt to simulate human auditory localization including the precedence effect. The algorithm is based on the well-known physiologically-motivated algorithm to estimate interaural time differences (ITDs), as originally proposed by Jeffress in Jeffres, L.A.: A place theory of sound localization, J. Comparative and Physiological Psychology, Vol. 41, 35-39, 1948. The Jeffress model consists of two delay lines, one from the left to the right ear, and the other one vice versa. Both are connected by several coincidence detectors, driving a central processor. A signal arriving at the left wear L(m), where m is the index for the time, has to pass a first delay unit 1(m,n), where n is the index for the coincidence detectors at different internal delays, to go from left to right along the delay line. In the same way, a signal arriving at the right ear R(m) travels on the other delay line r(m,n) into the opposite direction. The discrete implementation of the delay lines can be revealed mathematically as follows - with N being the number of the implemented coincidence cells.

$$1(m+1,n+1) = 1(m,n);\ 1 \le n < N \wedge 1(m,1) = L(m), \tag{1}$$

$$r(m+1,n-1) = r(m,n);\ 1 < n \le N \wedge r(m,N) = R(m), \tag{2}$$

[0030] A coincidence detector c(m,n) is activated when it receives simultaneous inputs from both delay lines, namely at the positions that it is connected to. Each of the coincidence detectors is adjusted to a different ITD, due to the limited velocity of propagation of the signals on the delay lines. For example, a sound source located in the left hemisphere will arrive at the left ear first. Such, the signal can travel a longer distance on the delay line than the signal on the delay line for the right ear - before both of them activate the coincidence detector for the corresponding ITD.

[0031] The probability for two spikes from the two opposite channels to coincide at a specific cell, tuned to a specific delay, is given by the product of the number of spikes in the left and right channels that pass one of these delay cells. This product also appears in the cross-correlation function, which is defined for a discrete system as follows:

$$\Psi_{L,R}(m,n)=\frac{1}{\Delta m}\sum_{m'=m}^{m+\Delta m}c(m',n)$$

(3)

[0032] with c(m',n) = l(m',n)r(m',n) and the assumption that the amplitude in the left and right channel is proportional to the number of spikes. It should be noted, that in this example a rectangular window duration Δm is used within which the cross-correlation function for each time interval is calculated.

[0033] Lindemann introduced contra-lateral inhibition elements to enhance this algorithm to sharpen the peaks in the binaural activity pattern and to suppress side-lobs which occur because of the periodic nature of the cross-correlation algorithm for band-pass signals. As a result other positions than those where the coincidence occurred are suppressed. Any activity running down at one of the two tapped delay lines can inhibit the activity at the corresponding positions on the opposite delay line. The implementation of the inhibition elements is achieved by modifying the computation of the delay lines (see equations 1-2):

$$1(m,n+1) = 1(m,n)\ [1-c_s \cdot 1(m,n)] \qquad (4)$$

$$r(m,n-1) = r(m,n)\ [1-c_s \cdot 1(m,n)] \qquad (4)$$

with static inhibition constant $0 \leq c_s < 1$.

[0034] A second objective of introducing (static) contra-lateral inhibition was to combine the influences of the ITDs and the interaural level differences (ILDs). The latter are of particular importance in the frequency range above 1.5 kHz. Thus, the results of the cross-correlation algorithm become dependent on the ILDs.

[0035] As the displacement due to ILD appeared to be too small, Lindemann introduced additional monaural processors into his model. For large ILDs not just the simple cross-correlation product c(m,n) = 1(m,n)r(m,n) is estimated, but rather the following equation:

$$c(n,m) = [(1-w(n))\cdot 1(n,m) + (1-w(n))\cdot r(n,m)]...$$

$$... [(1-w(N+1-n))\cdot 1(n,m) + (1-w(N+1-))\cdot 1(n,m)]\ (6) \qquad (6)$$

with $w(n) = 0.035 \cdot e^{-n/6}$, the monaural weighting function. For a general overview of the cross-correlation and inhibition structure see Fig. 5. Note, that monaural activity can be inhibited by binaural one, but not vice versa.

[0036] By using the cross-correlation algorithm, not only the lateral position of the cross-correlation peak, but also its width can be exploited to gain information on the auditory spaciousness of the room. There is a tendency that spaciousness increases with increasing decorrelation of the ear-input signals. This means that the broader the peaks in the correlation function are, the more spaciousness can be expected.

[0037] Gaik (Gaik, W. (1993): Combined evaluation of interaural time and intensity differences: Psychoacoustic results and computer modeling, J. Acoust. Soc. Am., Vol. 94, 98-110) introduced an enhancement to simulate the localization in the horizontal plane in a more exact way. He extended the Lindemann model to optimally process the natural combinations of ITDs and ILDs, as they are found in head-related transfer functions. For this purpose, Gaik introduced weighting factors into the delay line in such a way, that the ILD of the signal is compensated for before the two binaural signals meet at the corresponding coincidence detector. The "corresponding" detector in each frequency band is that one which represents the specific interaural delay which comes together with a specific ILD in a given HRTF. To be able to thus compensate the ILDs for various directions, proper weighting or trading factors have to be implemented between every two coincidence detectors. This is performed with a supervised learning procedure. Without Gaik's modification, the contra-lateral inhibition would bounce the cross-correlation peak sideways to unnatural positions, even with the applied ITDs and ILDs stemming from natural pairs of HRTFs. In the last processing step, the ITDs simply have to be remapped on the basis of the azimuth, to estimate the azimuth angle of the sound source.

[0038] It has to be mentioned at this point that, before the binaural cues can be analyzed adequately, the signals have to be processed by a stage that simulates the auditory periphery, namely, outer, middle and the inner ear (cochlea).

[0039] In Fig. 6 the different steps for determining the exact position comprising lateral deviation, time delay and sound level are shown in more detail. In a first step 610 the sound signal is generated, the signal being an impulse. The signal can be generated by the sound carte of the computer unit 14 or by any other device. This impulse is emitted from one of the loudspeakers 11 and the sound signal is detected in step 620 by the two microphones 13 in the left and the right ear of the dummy head. The sound signal is then processed in step 630 in accordance with the model shown in Fig. 4 and Fig. 5 and the deviation, time delay and/or sound level can be deduced from the binaural activity pattern in step 640.

[0040] In Fig. 7 the signal processing is shown in more detail. After the detection of the sound in step 620 the sound signal detected in the right and the left ear is further processed. In order to simulate the frequency decomposition as performed by the cochlea the left and the right ear signal are processed by a gammatone-filter bank 42 in step 631. Note, that the filtering by the outer ears is already encoded in the measured head-related impulse responses. To mimic the firing behavior of the inner hair cell response into account simple half-wave rectification is implemented in step 632. To consider that the fine structure of the rectified signals cannot be resolved in the higher frequency bands the signals are low-pass filtered after the half-wave rectification in step 633. After this, a running cross-correlation analysis comparing the signal of the left and the right channel was carried out in each frequency band (step 634). To consider the sluggishness of the human binaural system the temporal window and the running cross-correlation analysis was set to a duration of 10 ms. In step 635 the inhibition function is introduced. In the steps 634 and 635 the ITDs and the ILDs are calculated. The outputs of each frequency band are transformed from an ITD-scale to a scale representing the position on the right-left deviation scale in step 636. This scale represents one of the three coordinates of the BAP to be rendered. The time delay can be calculated from the intensity of the sound signal in dependence of time, the lateral deviation can be deduced from the intensity of the sound signal in dependence of the lateral position of the sound signal relative to the reference position, and the sound level can be deduced from the maximum of the sound signal. These three parameters determine the exact position of the sound source. In the case of a multi channel audio system these parameters could be output on a display unit of the computer unit 14, so that the user of the audio system knows whether the sound sources (loudspeakers) are positioned at the position prescribed by a standard. The user can then manually relocate the loudspeakers and run the above-described steps for determining the position again.

[0041] If the positioning of the loudspeakers at positions prescribed by the standards is not possible or is not desired these three parameters can also be input into a manipulation unit 16. This manipulation unit 16 will manipulate the sound emitted by the sound source for compensating the mismatch of the actual position of the different sound sources relative to the reference position of each sound source. After this adjustment the listener will have an optimal surround sound when positioned in the reference listening position, even though the loudspeakers are not positioned in the prescribed configuration.

[0042] In conclusion, it is possible to determine the position of a sound source using a simulation of the human auditory signal processing and by using the obtained binaural activity patterns. An automatic evaluation of the position is possible facilitating the setup of complex multi channel audio or audio and video systems.

Claims

1. Method for determining the position of a sound source (11) in relation to a reference position, comprising the steps of:

    - generating a sound signal emitted from the sound source (11),
    - detecting the emitted sound signal,
    - processing the sound signal by the use of a physiological model of the ear,
    - deducing at least one of lateral deviation in relation to the reference position,time delay of the sound signal from the sound source to the reference position, and the sound level of the detected sound signal.

2. Method according to claim 1, wherein the sound signal is detected by a sound receiving unit having two microphones (13) arranged spatially apart, resulting in two audio channels which are processed by the physiological model.

3. Method according to claim 2, wherein the two microphones (13) are positioned in the right and left ear of a dummy with replication of the external ear.

4. Method according to any of the preceding claims, wherein the sound signal is an impulse, and the head related, binaural impulse responses are detected.

5. Method according to any of the preceding claims, wherein the detected sound signal is processed by applying a gammatone filter bank (42) to the recorded signal.

6. Method according to any of the preceding claims, wherein the sound signal is processed by a half wave rectification (43) of the recorded sound signal.

7. Method according to any of the preceding claims, wherein the sound signal is processed by low pass filtering (44) the recorded sound signal.

8. Method according to any of claims 2 to 7, wherein the two channel audio signal is processed by carrying out a cross correlation analysis of the sound signals of the left and the right ear, said cross correlation analysis resulting in the Interaural Time Differences (ITD) of the left and right ear.

9. Method according to any of the preceding claims, wherein the signal is processed by introducing inhibition elements, resulting in the Interaural Level Differences (ILD) of the left and right ear.

10. Method according to claim 8 and 9, wherein the signal processing comprises the step of deducing a Binaural Activity Pattern (BAP) from the Interaural time differences (ITD) and the Interaural Level Differences (ILD), the binaural activity pattern comprising the intensity of the sound signal in dependence of time and comprising the intensity of the sound signal in dependence of the lateral deviation of the sound source (11).

11. Method according to claim 10, wherein the time delay can be deduced from the intensity of the sound signal in dependence of time, the lateral deviation being deduced from the intensity of the sound signal in dependence of the lateral position of the sound signal relative to the reference position, the sound level being deduced from the maximum of the sound signal.

12. Method according to claim 11, wherein the lateral deviation, the time delay and the sound level can be used in order to calculate correction data which adjust the sound signal emitted from the sound sources according to an arrangement of sound sources (11) which correspond to a predetermined standard arrangement of the sound sources.

13. Method according to claim 11, wherein the predetermined standard arrangement of the sound sources corresponds to the arrangement of the sound sources according to an international standard, such as the ITU-R BS 775-1.

14. Method according to any of the preceding claims, wherein the individual positions of a plurality of sound sources (11) of a multichannel audio system are determined.

15. System (10) for determining the position of a sound source, comprising

- a sound generator for generating sound
- a sound source (11) for emitting the generated sound,
- a sound receiving unit (13) for detecting the sound emitted from the sound sources, and
- sound analysing means (15) for processing the detected sound signal by the use of a physiological model of the ear and for deducing at least one of lateral deviation in relation to a reference position,time delay of the sound signal from the sound source (11) to the reference position, and the sound level of the detected sound signal.

16. System according to 15, **characterized in that** the sound receiving unit comprises two microphones which are arranged spatially apart.

17. System according to claim 16, further comprising a dummy (12) with two ears in which the microphones are positioned, the ears comprising replications of the external human ear.

18. System according to any of claims 15 to 17, **characterized in that** the sound analysing means comprises a gammatome filter bank (42).

19. System according to any of claims 15 to 18, **characterized in that** the sound analysing means comprises a half-wave rectification unit (43).

20. System according to any of claims 15 to 19, **characterized in that** the sound analysing means comprises a low pass filter (44).

21. System according to any of claims 15 to 20, **characterized in that** the sound analysing means comprises a cross correlation analysis unit (45).

22. System according to any of claims 15 to 21, **characterized by** further comprising a manipulation unit (16) for manipulating the sound emitted by the sound source for compensating a mismatch of actual position of the different sound sources relative to the reference position of each sound source.

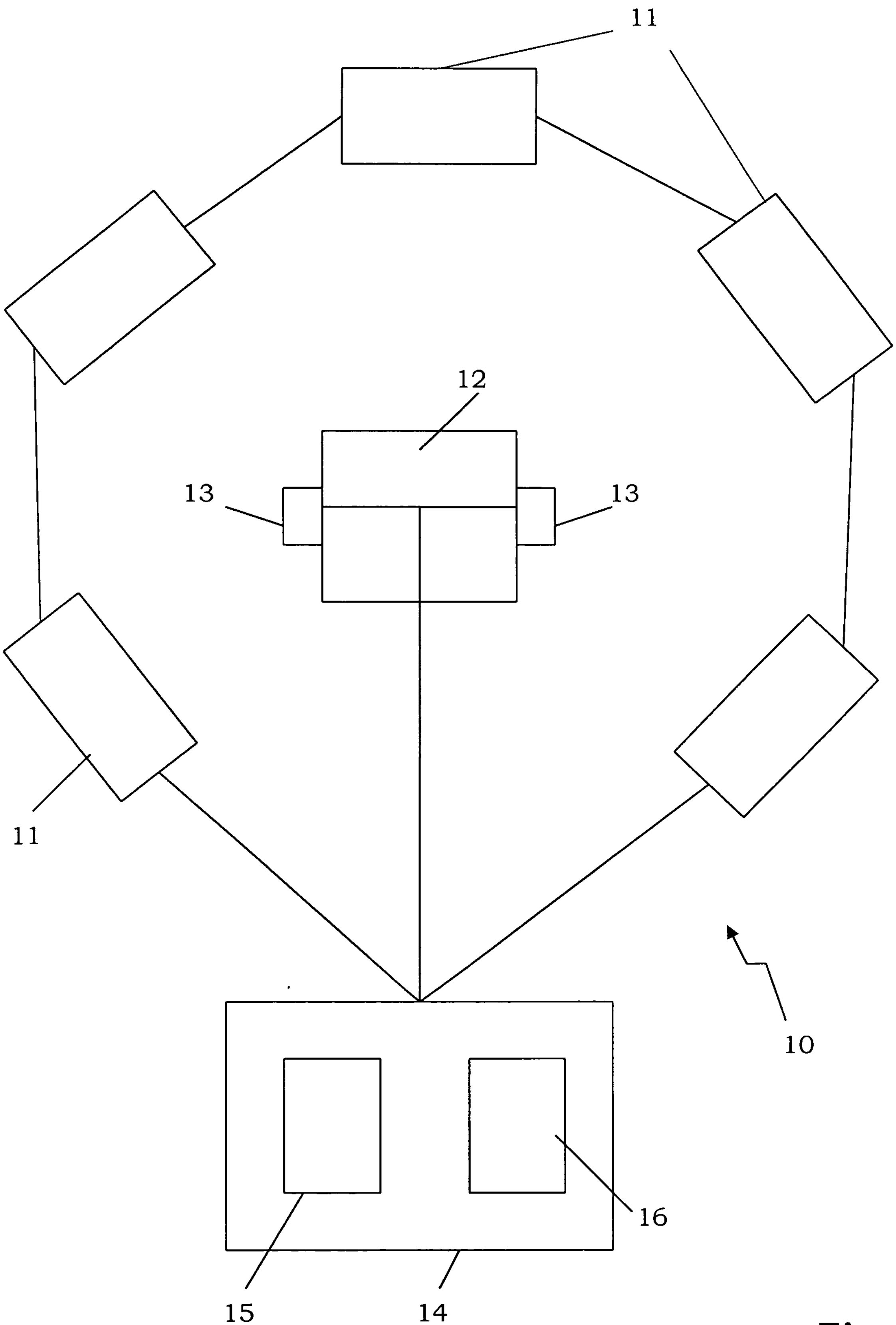


Fig. 1

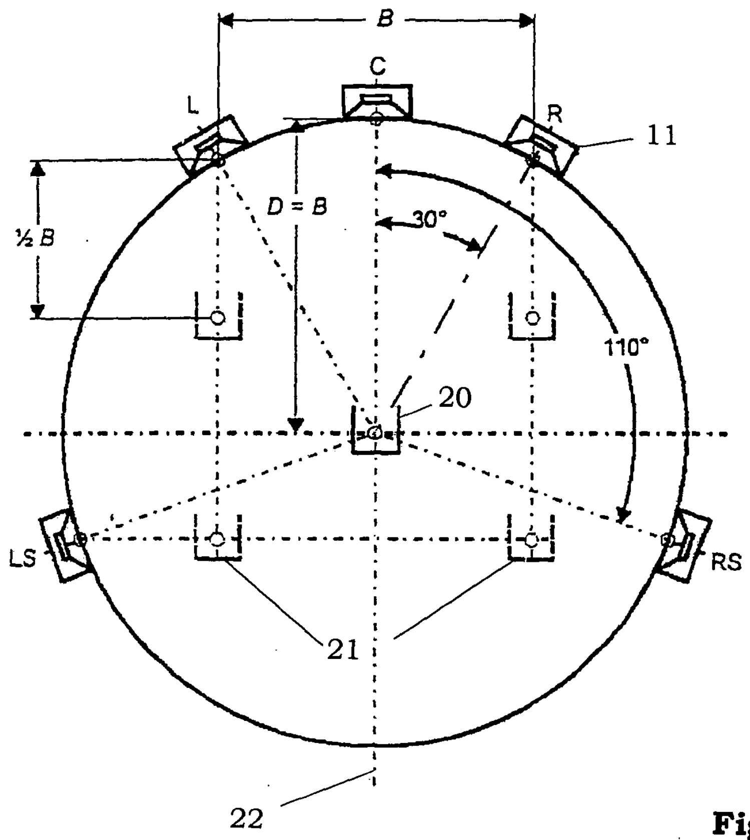


Fig. 2

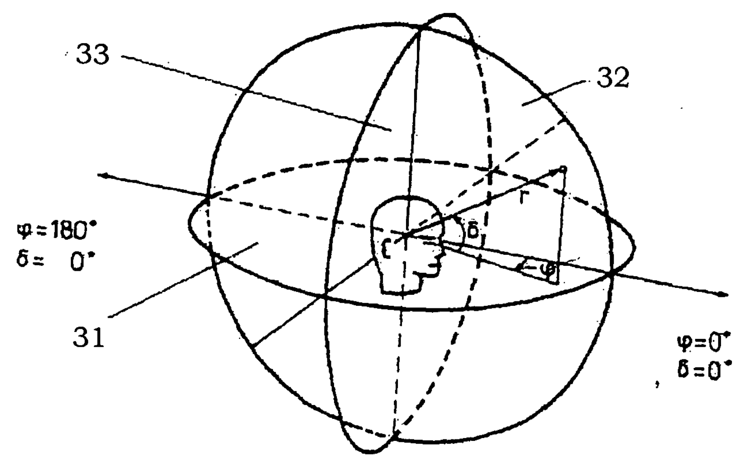
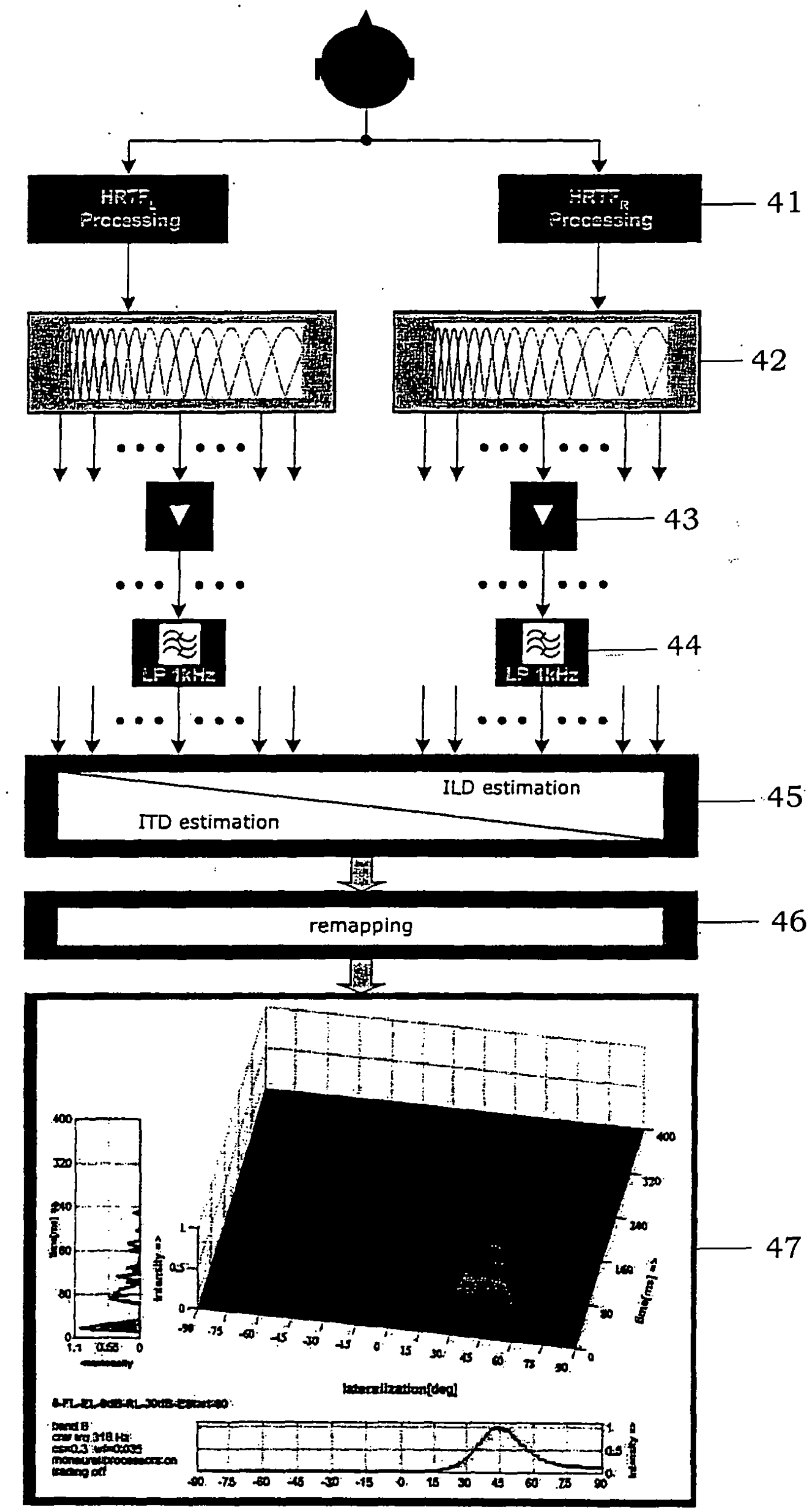


Fig. 3



HRTF$_L$ Processing
HRTF$_R$ Processing
41
42
43
LP 1kHz
LP 1kHz
44
ILD estimation
ITD estimation
45
remapping
46
lateralization[deg]
band 8
47

Fig. 4

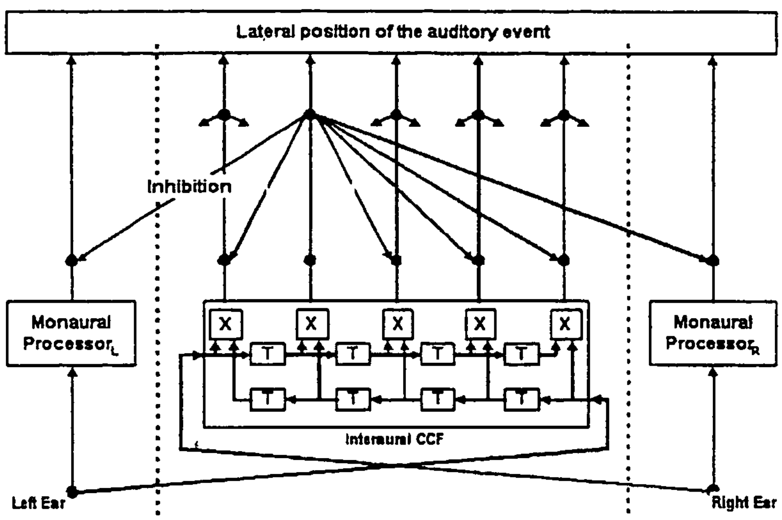

Lateral position of the auditory event
Inhibition
Monaural Processor$_L$
X
T
Interaural CCF
Monaural Processor$_R$
Left Ear
Right Ear

Fig. 5

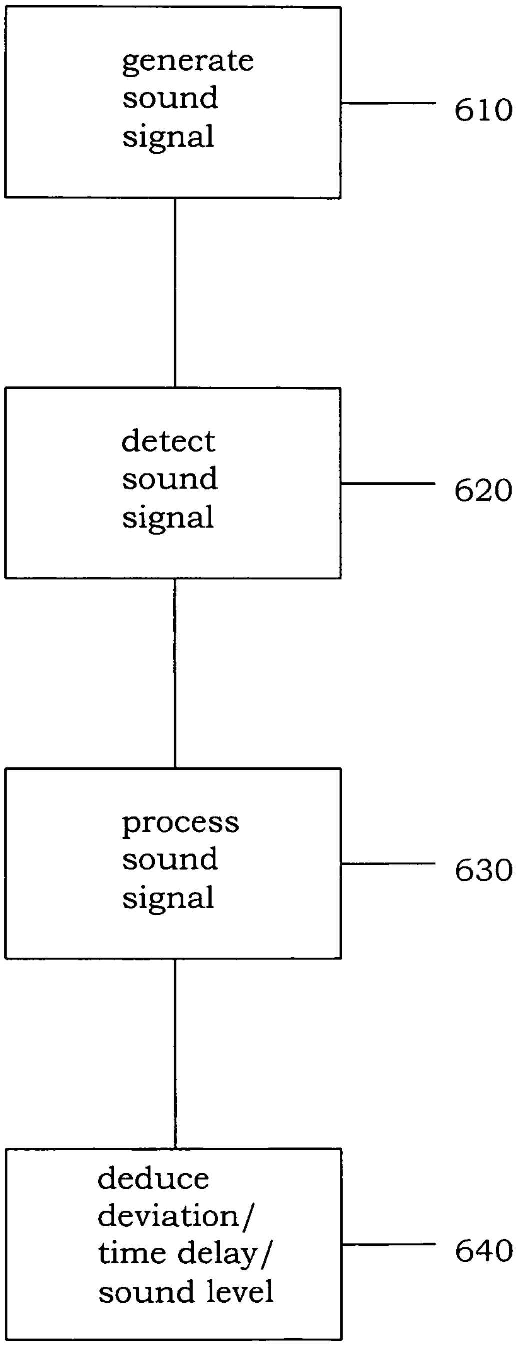

generate sound signal
610
detect sound signal
620
process sound signal
630
deduce deviation/ time delay/ sound level
640

Fig. 6

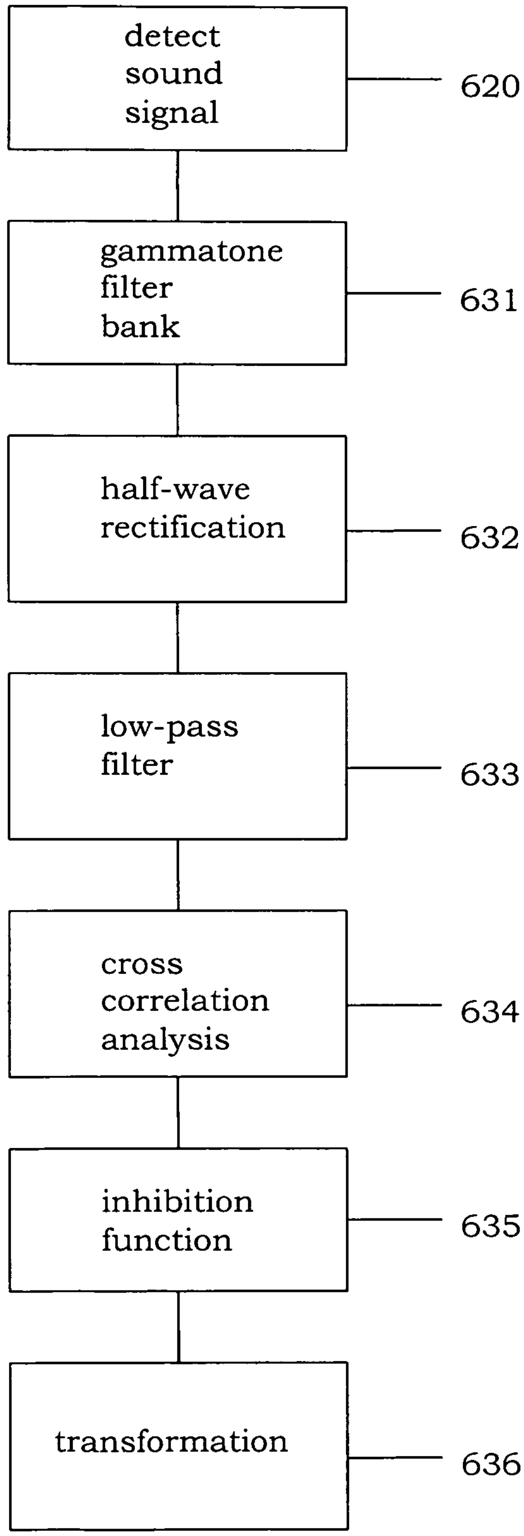

detect sound signal
620
gammatone filter bank
631
half-wave rectification
632
low-pass filter
633
cross correlation analysis
634
inhibition function
635
transformation
636

Fig. 7

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 03 02 2983

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.7)
X	CHIANG-JUNG PU ET AL: "A neuromorphic microphone for sound localization" SYSTEMS, MAN, AND CYBERNETICS, 1997. COMPUTATIONAL CYBERNETICS AND SIMULATION., 1997 IEEE INTERNATIONAL CONFERENCE ON ORLANDO, FL, USA 12-15 OCT. 1997, NEW YORK, NY, USA,IEEE, US, 12 October 1997 (1997-10-12), pages 1469-1474, XP010249534 ISBN: 0-7803-4053-1	1,4-7, 15,18-20	G01S5/18 G01S3/808
Y	* page 1469 - page 1472 *	2,3,16, 17	

X	NAKADAI K ET AL: "Auditory Fovea based speech separation and its application to dialog system" PROCEEDINGS OF THE 2002 IEEE/RSJ INTERNATIONAL CONFERENCE ON INTELLIGENT ROBOTS AND SYSTEMS. (IROS 2002). LAUSANNE, SWITZERLAND, SEPT. 30 - OCT. 4, 2002, IEEE/RSJ INTERNATIONAL CONFERENCE ON INTELLIGENT ROBOTS AND SYSTEMS, NEW YORK, NY: IEEE, US, vol. 1 OF 3, 30 September 2002 (2002-09-30), pages 1320-1325, XP010609601 ISBN: 0-7803-7398-7 * page 1321 - page 1322 *	1,15	TECHNICAL FIELDS SEARCHED (Int.Cl.7) G01S

	-/--		

The present search report has been drawn up for all claims

1

Place of search	Date of completion of the search	Examiner
THE HAGUE	26 February 2004	Devine, J

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons
& : member of the same patent family, corresponding document

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 03 02 2983

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.7)
Y	GEORGIOU P G ET AL: "A multiple input single output model for rendering virtual sound sources in real time" MULTIMEDIA AND EXPO, 2000. ICME 2000. 2000 IEEE INTERNATIONAL CONFERENCE ON NEW YORK, NY, USA 30 JULY-2 AUG. 2000, PISCATAWAY, NJ, USA,IEEE, US, 30 July 2000 (2000-07-30), pages 253-256, XP010511447 ISBN: 0-7803-6536-4 * page 254 - page 255 * ---	2,3,16, 17	
A	CHANG P-R ET AL: "Fuzzy neural systems for controlling sound localisation in stereophonic reproduction" IEE PROCEEDINGS D. CONTROL THEORY & APPLICATIONS, INSTITUTION OF ELECTRICAL ENGINEERS. STEVENAGE, GB, vol. 145, no. 4, 28 July 1998 (1998-07-28), pages 393-401, XP006010993 ISSN: 1350-2379 * page 393 - page 394 * ---	1,15	TECHNICAL FIELDS SEARCHED (Int.Cl.7)
A	US 5 381 512 A (HOLTON THOMAS ET AL) 10 January 1995 (1995-01-10) * column 22 - column 25 * -----	6	

1 The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	26 February 2004	Devine, J

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons
& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

ANNEX TO THE EUROPEAN SEARCH REPORT ON EUROPEAN PATENT APPLICATION NO. EP 03 02 2983

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report.
The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

26-02-2004

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5381512 A	10-01-1995	NONE	

EPO FORM P0459

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82